Subject-level Joint Parcellation-Detection-Estimation in fMRI

Lotfi Chaari, Solveig Badillo, Thomas Vincent, Ghislaine Dehaene-Lambertz, Florence Forbes, Philippe Ciuciu

Subject-level Joint Parcellation-Detection-Estimation in fMRI

Lotfi CHAARI *Member*, Solveig BADILLO, Thomas VINCENT, Ghislaine DEHAENE-LAMBERTZ, Florence FORBES and Philippe CIUCIU *Senior Member*

***Abstract*—Brain parcellation is one of the most important issues in functional MRI (fMRI) data analysis. This parcellation allows establishing homogeneous territories that share the same functional properties. This paper presents a model-based approach to perform a subject-level parcellation into hemodynamic territories with similar hemodynamic features which are known to vary between brain regions. We specifically investigate the use of the Joint Parcellation-Detection-Estimation (JPDE) model initially proposed in [1] to separate brain regions that match different hemodynamic response function (HRF) profiles. A hierarchical Bayesian model is built and a variational expectation maximization (VEM) algorithm is deployed to perform inference. A more complete version of the JPDE model is detailed. Validation on synthetic data shows the robustness of this model to varying signal-to-noise ratio (SNR) as well as to different initializations. Our results also demonstrate that good parcellation performance is achieved even though the parcels do not involve the same amount of activation. On real fMRI data acquired in children during a language paradigm, we retrieved a parcellation along the superior temporal sulcus of the left hemisphere that matches the gradient of activation dynamics already reported in the literature.**

I. Introduction

Uncovering the functional organization of the brain from blood oxygen level dependent (BOLD) fMRI data is one of the most exciting challenges in neuroimaging nowadays. A functional parcellation of the brain, which can be seen as a segmentation of fMRI data into homogeneous brain regions according to some specific features (eg, BOLD effect size), can help understand brain organization. The reason for this is at least twofold: on the one hand, the functional parcellation is a key segmentation feature that allows neuroscientists to perform group-level statistical inference at a coarser spatial resolution than the voxel one, hence compensating for inter-subject anatomo-functional variability and spatial normalisation errors [2]. On the other hand, grouping similar voxels or regions together from a functional viewpoint can be seen as a way of reducing the multiple comparison problem that systematically arise in whole brain statistical inference. Hence, functional parcellation has been also developed for improving reliability and reproducibility of group-level statistical fMRI studies [3]. Of note, even though these motivations originally emerge from task-related fMRI studies [2–7], the emergence of resting-state fMRI and functional connectivity studies (either based on pairwise correlation or on multivariate methods) [8–12] has also emphasized the need for segmenting the functional connectome (or brain networks) using parcellation techniques. This has recently appeared as the way of selecting the most common intrinsic brain activity to all individuals in a given cohort [13–15].

In this paper, in the context of task-related fMRI studies, we are rather interested in brain parcellation for uncovering homogeneous *hemodynamic territories* from fMRI data, which to the best of our knowledge has barely been addressed in the literature [1, 7, 16]. Ideally, this concern should be addressed at the population level but owing to the lack of existing hemodynamic brain atlas this work starts by proposing subject-level hemodynamic parcellation, leaving the group-level extension to future investigations. Identifying a fine hemodynamic parcellation in each individual is a challenging task given the existing between-region variability in the hemodynamic response function (HRF) [17, 18]. This challenge clearly calls for the modeling and estimation of HRF shapes that reflect the specificity of hemodynamic features at a parcel level [19]. Such an hemodynamic parcellation can be helpful to improve the detection of evoked activity and localize which regions of the brain are involved in task performance [20, 21]. Indeed, this allows to improve on the use of a single general linear model (GLM) for detecting evoked activity [22], which assumes a fixed and constant HRF shape for the whole brain. Instead, individually calibrated HRF shapes can be considered and the detection task performed with spatially varying GLM models [20].

To avoid any circularity issue between detection of brain activity and estimation of the underlying hemodynamics, both tasks have been confronted to one another and thus addressed jointly in the joint detection-estimation framework (JDE) [19, 20]. The JDE model has been developed at the parcel level which means that a single parcel specific unknown HRF shape is driving hemodynamic responses in a given parcel. Then, a spatially regularized bilinear model is inferred in each parcel to localize activated voxels. However, the JDE approach suffers from a major limitation: it relies on a fixed parcellation that needs to be defined a priori using for instance clustering algorithms. To avoid such a commitment to a pre-specified parcellation, a first JDE extension referred to as the joint parcellation-detection-estimation (JPDE) approach, has been proposed in [1]. The motivation for this extension was twofold: (i) to group the regions that share a similar HRF pattern and (ii) to relax the hard constraint of a single HRF profile over a given parcel to cope with possible parcellation errors. Both concerns were methodologically addressed by introducing HRF patterns, each of them being represented by a Gaussian distribution and assigned to its representative voxels through latent variables. To favor large connected parcels, these latent variables are governed by a hidden Markov model (Potts field) that allows the enforcement of spatial correlation. Using the additional parcellation layer introduced in the JPDE model, hemodynamics estimation hence reduces to the identification of a limited number (say K) of HRF patterns. The performance of JPDE has been illustrated on few simulated examples as well as on over-simplistic real fMRI data in [1]. In this paper, we first provide a complete development of the JPDE model with a fully detailed description of the different steps of the algorithm (expectation and maximization steps). In particular, we complete the original version of JPDE [1] by proposing an automated setting of the correlation parameter in the Potts field. This makes the JPDE model inference more computationally demanding but more user-friendly. We also investigate on simulations its robustness to a large number of confounding factors such as a varying signal-to-noise ratio (SNR), a fluctuating overlap between activated voxels and parcel definition, and a number of critical issues like different parcellation initializations. Last but not least, we then fit the JPDE model to more challenging real fMRI data acquired on a 6 year-old child and dedicated to uncover the functional and hemodynamic organization

A short version of this work has been published in [1]

L. CHAARI is with the University of Toulouse, IRIT - INP-ENSEEIHT (UMR 5505), 2 rue Charles Camichel, BP 7122, Toulouse Cedex 7 France. E-mail: lotfi.chaari@enseeiht.fr. T. VINCENT and F. FORBES are with the Mistis team at INRIA Grenoble, 655 avenue de l'Europe, Montbonnot, 38334 Saint Ismier Cedex, France. E-mail: {thomas.vincent,florence.forbes}@inria.fr. P. CIUCIU and G. DEHAENE-LAMBERTZ are with CEA/DSV/I^{2}BM/NeuroSpin, CEA Saclay, Bat. 145, Point Courrier 156, 91191 Gif-sur-Yvette cedex, France. E-mail: {philippe.ciuciu,Ghislaine.Dehaene-Lambertz}@cea.fr. P. CIUCIU is also with and INRIA Parietal team and G. DEHAENE-LAMBERTZ with INSERM U992, Cognitive neuroimaging Unit.

of the left superior temporal sulcus (STS) during a speech perception experiment.

The rest of this paper is organized as follows. The JPDE model is first introduced in Section II. The complete variational expectation maximization procedure is then fully described in Section III. Results on both artificial and real fMRI data are provided in Section IV. We discuss the pros and cons of this extended JPDE approach in Sections V and concluding remarks are drawn in Section VI.

II. A joint Parcellation-Detection-Estimation model

A. Observed and latent variables

We rely here on the same observation model as the one proposed in the original JPDE framework [1]. At voxel $j \in \boldsymbol{\mathcal{V}}$ ($\boldsymbol{\mathcal{V}}$ being the set of brain voxels), the fMRI time series $\boldsymbol{y}_j$ is measured at times $\{t_n, n = 1{:}N\}$, where $t_n = n\,TR$, N being the number of scans and TR the time of repetition. The whole set of time series is denoted by $\boldsymbol{Y} = \{\boldsymbol{y}_j, j \in \boldsymbol{\mathcal{V}}\}$. The number of different stimulus types or experimental conditions is M. At each voxel j, we assume a voxel dependent HRF $\boldsymbol{h}_j \in \mathbb{R}^{D+1}$ with $\boldsymbol{H} = \{\boldsymbol{h}_j, j \in \boldsymbol{\mathcal{V}}\}$ the set of all HRFs. Each $\boldsymbol{h}_j$ is associated with a specific HRF group among K, each group being represented by a specific HRF pattern. These groups are specified by a set of hidden labels $\boldsymbol{Z} = \{z_j, j \in \boldsymbol{\mathcal{V}}\}$ where $z_j \in \{1 : K\}$ and $z_j = k$ means that voxel j is representative of the k-th group. Inferring $\boldsymbol{Z}$ thus corresponds to segmenting the brain mask into K hemodynamic territories whose connected components define a parcellation. The link between observed BOLD data and missing variables is then specified by:

$$\boldsymbol{y}_j = \boldsymbol{S}_j\boldsymbol{h}_j + \boldsymbol{P}\boldsymbol{\ell}_j + \boldsymbol{b}_j, \quad \text{with} \quad \boldsymbol{S}_j = \sum_{m=1}^{M} a_j^m \boldsymbol{X}_m \tag{1}$$

where $\boldsymbol{S}_j\boldsymbol{h}_j$ is the sum over the stimulus-induced components in the BOLD signal. The same binary matrix $\boldsymbol{X}_m$ as in [1, 19–21] is used to encode the stimuli arrival times. Note here that the generative model in (1) is different from the one in [20, 21] since the HRF in the present paper is voxel-dependent ($\boldsymbol{h}_j$).

We write $\boldsymbol{A} = \{\boldsymbol{a}^m, m = 1 : M\}$ with $\boldsymbol{a}^m = \{a_j^m, j \in \boldsymbol{\mathcal{V}}\}$ where the scalar a_j^m's are voxel-dependent weights that model the transition between stimulations and the neuro-vascular response in response to the m-th stimulus type. They are generally referred to as Neural Response Levels (NRL). Similarly to the HRF's, each NRL is assumed to be in one of I groups specified by activation class assignment variables $\boldsymbol{Q} = \{\boldsymbol{q}^m, m = 1 : M\}$ where $\boldsymbol{q}^m = \{q_j^m, j \in \boldsymbol{\mathcal{V}}\}$ and q_j^m represents the *activation class* at voxel j for condition m. The number of classes considered here is $I = 2$ for non-activated ($i = 0$) and activated ($i = 1$) voxels. Finally, the rest of the signal is made of vector $\boldsymbol{P}\boldsymbol{\ell}_j$, which corresponds to low frequency drifts with $\boldsymbol{P}$ a $N \times O$ matrix, $\boldsymbol{\ell}_j \in \mathbb{R}^O$ a vector to be estimated and $\boldsymbol{L} = \{\boldsymbol{\ell}_j, j \in \boldsymbol{\mathcal{V}}\}$. Regarding the observation noise, the $\boldsymbol{b}_j$'s are assumed to be independent with $\boldsymbol{b}_j \sim \mathcal{N}(0, \boldsymbol{\Gamma}_j^{-1})$ meaning that $\boldsymbol{b}_j$ follows a Gaussian distribution with mean 0 and covariance matrix $\boldsymbol{\Gamma}_j^{-1}$. The set of all unknown precision matrices is denoted by $\boldsymbol{\Gamma} = \{\boldsymbol{\Gamma}_j, j \in \boldsymbol{\mathcal{V}}\}$.

B. Hierarchical model of the complete data distribution

Additional standard assumptions [19–21, 23] lead to the model: $p(\boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{H}, \boldsymbol{Q}, \boldsymbol{Z}) = p(\boldsymbol{Y} \mid \boldsymbol{A}, \boldsymbol{H})\; p(\boldsymbol{A} \mid \boldsymbol{Q})\; p(\boldsymbol{Q})\; p(\boldsymbol{H} \mid \boldsymbol{Z})\; p(\boldsymbol{Z})$. In the above expression, parameters have not been specified for simplicity. Each term in the product is then defined as follows.

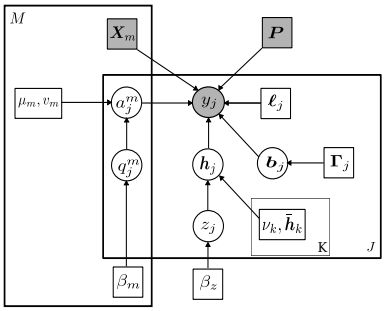

Fig. 1. Graphical model describing dependencies between *observed* and *latent* variables involved in the JPDE model with J voxels. Circles and squares indicate random variables and model parameters, respectively. Observed variables and fixed parameters are shaded. We used standard graphical notations where plates represent multiple similar nodes with their number given in the plate corner.

1) Likelihood:

Akin to [19, 21, 23, 24], an autoregressive (AR) noise model has been adopted to account for serial correlations in fMRI time series. We will assume a noise model $\boldsymbol{b}_j \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Gamma}_j^{-1})$ with $\boldsymbol{\Gamma}_j = \sigma_j^{-2}\boldsymbol{\Lambda}_j$ where $\boldsymbol{\Lambda}_j$ is a tridiagonal symmetric matrix which depends on the AR(1) parameter ρ_j [19]: $(\boldsymbol{\Lambda}_j)_{1,1} = (\boldsymbol{\Lambda}_j)_{N,N} = 1$, $(\boldsymbol{\Lambda}_j)_{n,n} = 1 + \rho_j^2$ for $n \in \{2 : (N-1)\}$ and $(\boldsymbol{\Lambda}_j)_{n+1,n} = (\boldsymbol{\Lambda}_j)_{n,n+1} = -\rho_j$ for $n \in \{1 : (N-1)\}$. These parameters are assumed voxel-specific due to their tissue-dependence [25, 26]. Denoting $\overline{\boldsymbol{y}}_j = \boldsymbol{y}_j - \boldsymbol{P}\boldsymbol{\ell}_j - \boldsymbol{S}_j\boldsymbol{h}_j$, the likelihood can be factorized over voxels as follows:

$$p(\boldsymbol{Y} \mid \boldsymbol{A}, \boldsymbol{H}; \boldsymbol{L}, \boldsymbol{\Gamma}) \propto \prod_{j \in \boldsymbol{\mathcal{V}}} \det\boldsymbol{\Lambda}_{\mathrm{j}}^{1/2} \sigma_{\mathrm{j}}^{-\mathrm{N}} \exp\Big(-\frac{\overline{\boldsymbol{y}}_{\mathrm{j}}^{\mathrm{t}} \boldsymbol{\Lambda}_{\mathrm{j}} \overline{\boldsymbol{y}}_{\mathrm{j}}}{2\sigma_{\mathrm{j}}^2}\Big). \tag{2}$$

2) Neural response levels:

The NRLs are assumed to be statistically independent across conditions: $p(\boldsymbol{A}; \boldsymbol{\theta}_{\boldsymbol{a}}) = \prod_{m=1}^{M} p(\boldsymbol{a}^m; \boldsymbol{\theta}_m)$ where $\boldsymbol{\theta}_{\boldsymbol{a}} = \{\boldsymbol{\theta}_m, m = 1 : M\}$ and $\boldsymbol{\theta}_m$ gathers the parameters for the m-th condition. A mixture model is then adopted by using the latent variables q_j^m to segregate non-activated voxels ($q_j^m = 0$) from activated ones ($q_j^m = 1$). For the m-th condition, and conditionally to the assignment variables $\boldsymbol{q}^m$, the NRLs are assumed to be independent: $p(\boldsymbol{a}^m \mid \boldsymbol{q}^m; \boldsymbol{\theta}_m) = \prod_{j \in \boldsymbol{\mathcal{V}}} p(a_j^m \mid q_j^m; \boldsymbol{\theta}_m)$ with $p(a_j^m \mid q_j^m = i; \boldsymbol{\theta}_m) \sim \mathcal{N}(\mu_{mi}, v_{mi})$ and $\boldsymbol{\theta}_m = \{\mu_{mi}, v_{mi}, i = 0, 1\}$. We also denote $\boldsymbol{\mu} = \{\boldsymbol{\mu}_{mi}, m = 1 : M, i = 0, 1\}$ and $\boldsymbol{v} = \{\boldsymbol{v}_{mi}, m = 1 : M, i = 0, 1\}$. For non-activating voxels ($i = 0$) we set $\mu_{m0} = 0$, for all $m = 1 : M$. The other parameters are unknown and have to be estimated.

3) Activation classes:

As in [20], we assume prior independence between the M experimental conditions regarding the activation class assignments. It follows that $p(\boldsymbol{Q}) = \prod_{m=1}^{M} p(\boldsymbol{q}^m; \beta_m)$ where we assume in addition that $p(\boldsymbol{q}^m; \beta_m)$ is a Markov random field, namely a 2-class Potts model. Such prior modeling assumption is consistent with the physiological properties of the fMRI signal where the activity is known to be correlated in space [25, 27]. Here, the prior Potts model with interaction parameter β_m [20] is expressed, omitting the intractable normalizing constant, as:

$$p(\boldsymbol{q}^m; \beta_m) \propto \exp\big(\beta_m U(\boldsymbol{q}^m)\big) \tag{3}$$

with $U(\boldsymbol{q}^m) = \sum_{(j,j') \in \boldsymbol{\mathcal{V}}^2, j \sim j'} I(q_j^m = q_{j'}^m)$ and where for all $(a, b) \in \mathbb{R}^2$, $I(a = b) = 1$ if $a = b$ and 0 otherwise. The notation $j \sim j'$ means that the summation is over all neighboring voxels. The neighboring system may cover a 3D scheme through

the brain volume. The unknown parameters are denoted by $\boldsymbol{\beta} = \{\beta_m, m = 1 : M\}$. In what follows, we will consider a 6-connexity 3D neighboring system.

4) HRF groups: In order to promote large connected parcels, we introduce here an additional Markov random field, namely a K-class Potts model with interaction parameter β_z:

$$p(\boldsymbol{Z}; \beta_z) \propto \exp\big(\beta_z U(\boldsymbol{Z})\big) \tag{4}$$

where $U(\boldsymbol{Z}) = \sum_{(j,j') \in \boldsymbol{\mathcal{V}}^2, j \sim j'} I(z_j = z_{j'})$. It results from the above model that neighboring voxels tend to belong to the same HRF group, sharing thus the same HRF pattern.

5) HRF patterns: In contrast to [19–21] where a unique HRF shape is considered for a whole parcel, the distribution of $\boldsymbol{h}_j$ is expressed, for each voxel j, conditionally to the HRF group variable z_j: $p(\boldsymbol{H}|\boldsymbol{Z}) = \prod_{j \in \boldsymbol{\mathcal{V}}} p(\boldsymbol{h}_j \,|\, z_j)$ with $p(\boldsymbol{h}_j \,|\, z_j = k) \sim \mathcal{N}(\bar{\boldsymbol{h}}_k, \bar{\boldsymbol{\Sigma}}_k)$. Here, the mean vector $\bar{\boldsymbol{h}}_k$ can be seen as the HRF pattern for group k and $\bar{\boldsymbol{\Sigma}}_k$ regulates the stochastic perturbations around $\bar{\boldsymbol{h}}_k$. In practice, we will consider $\bar{\boldsymbol{\Sigma}}_k = \nu_k I_{(D+1)}$. In addition, smooth $\bar{\boldsymbol{h}}_k$'s are favored by controlling their second order derivatives with the following prior: $\bar{\boldsymbol{h}}_k \sim \mathcal{N}(\boldsymbol{0}, \sigma_h^2 \boldsymbol{R})$ with $\boldsymbol{R} = (\Delta t)^4 \ (\boldsymbol{D}_2^{\mathrm{t}} \boldsymbol{D}_2)^{-1}$ where $\boldsymbol{D}_2$ is the second-order finite difference matrix and σ_h^2 is a hyperparameter to be estimated or fixed. Moreover, $\bar{\boldsymbol{h}}_{k0} = \bar{\boldsymbol{h}}_{kD\Delta t} = 0$ as in [19–21]. Using the notation $\bar{\boldsymbol{h}} = \big\{\bar{\boldsymbol{h}}_k, k = 1 : K\big\}$ and $\boldsymbol{\nu} = \{\nu_k, k = 1 : K\}$, the parameters are denoted by $\boldsymbol{\Theta} = \big\{\boldsymbol{\Gamma}, \boldsymbol{L}, \boldsymbol{\mu}, \boldsymbol{v}, \boldsymbol{\beta}, \beta_z, \sigma_h^2, \boldsymbol{\nu}, \bar{\boldsymbol{h}}\big\}$ and belong to a set $\underline{\boldsymbol{\Theta}}$. In the following developments, the $\bar{\boldsymbol{h}}_k$'s will be considered as random parameters following the prior specified above with hyperparameter σ_h^2.

III. Variational EM estimation

We propose to use an EM framework to deal with the missing data $\boldsymbol{A} \in \mathcal{A}$, $\boldsymbol{H} \in \mathcal{H}$, $\boldsymbol{Q} \in \mathcal{Q}$, $\boldsymbol{Z} \in \mathcal{Z}$. From a methodological viewpoint, the adopted approach will be similar to the one in [1, 21, 23].

A. Variational Expectation-Maximization

Let $\mathcal{D}$ be the set of all probability distributions on $\mathcal{A} \times \mathcal{H} \times \mathcal{Q} \times \mathcal{Z}$. EM can be viewed [28] as an alternating maximization procedure of the free energy $\mathcal{F}$ on $\mathcal{D}$, for all $\tilde{p} \in \mathcal{D}$,

$$\mathcal{F}(\tilde{p}, \boldsymbol{\Theta}) \quad = \quad \mathrm{E}_{\tilde{p}}\big[\log p(\boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{H}, \boldsymbol{Q}, \boldsymbol{Z} \,;\, \boldsymbol{\Theta})\big] + \mathcal{G}(\tilde{p}) \tag{5}$$

where $\mathrm{E}_{\tilde{p}}\big[.\big]$ denotes the expectation with respect to $\tilde{p}$ and $\mathcal{G}(\tilde{p}) = -\mathrm{E}_{\tilde{p}}\big[\log \tilde{p}(\boldsymbol{A}, \boldsymbol{H}, \boldsymbol{Q}, \boldsymbol{Z})\big]$ is the entropy of $\tilde{p}$. Maximizing this free energy with respect to $\tilde{p}$ amounts to minimizing the Kullback-Leibler divergence between $\tilde{p}$ and the posterior distribution of interest $p(\boldsymbol{A}, \boldsymbol{H}, \boldsymbol{Q}, \boldsymbol{Z} \,|\, \boldsymbol{Y}; \boldsymbol{\Theta})$. The alternating procedure of the EM algorithm at iteration (r) and by denoting the current parameter values by $\boldsymbol{\Theta}^{(r-1)}$ writes:

$$\textbf{E-step: } \tilde{p}_{A,H,Q,Z}^{(r)} = \arg\max_{\tilde{p} \in \mathcal{D}} \ \mathcal{F}(\tilde{p}, \boldsymbol{\Theta}^{(r-1)}) \tag{6}$$

$$\textbf{M-step: } \boldsymbol{\Theta}^{(r)} = \arg\max_{\Theta \in \underline{\boldsymbol{\Theta}}} \ \mathcal{F}(\tilde{p}_{A,H,Q,Z}^{(r)}, \boldsymbol{\Theta}) \tag{7}$$

As reported in [28], the optimization step in Eq. (6) leads to $\tilde{p}_{A,H,Q,Z}^{(r)} = p(\boldsymbol{A}, \boldsymbol{H}, \boldsymbol{Q}, \boldsymbol{Z} \,|\, \boldsymbol{Y}; \boldsymbol{\Theta}^{(r-1)})$ which is intractable. A variational approximation (resulting in a variational EM algorithm: VEM) is therefore used to optimize over the distributions in $\mathcal{D}$ that factorize as a product of four pdfs on $\mathcal{A}$, $\mathcal{H}$, $\mathcal{Q}$ and $\mathcal{Z}$, respectively.

B. Variational Joint Parcellation-Detection-Estimation

We resort to the same VEM procedure as in [1]. At each iteration (r), we approximate the intractable posterior $p(\boldsymbol{A}, \boldsymbol{H}, \boldsymbol{Q}, \boldsymbol{Z} \,|\, \boldsymbol{Y}; \boldsymbol{\Theta}^{(r-1)})$ by a product of pdfs $\widetilde{p}_H^{(r)}$, $\widetilde{p}_A^{(r)}$, $\widetilde{p}_Q^{(r)}$ and $\widetilde{p}_Z^{(r)}$.

Our E-step is therefore approximated using a decomposition in four sub-steps consisting of alternatively updating $\widetilde{p}_H^{(r)}$, $\widetilde{p}_A^{(r)}$, $\widetilde{p}_Q^{(r)}$ and $\widetilde{p}_Z^{(r)}$. At iteration (r) with current estimates denoted by $\widetilde{p}_H^{(r-1)}, \widetilde{p}_A^{(r-1)}, \widetilde{p}_Q^{(r-1)}$, $\widetilde{p}_Z^{(r-1)}$ and $\boldsymbol{\Theta}^{(r-1)}$, the updating rules become:

$$\textbf{VE-H: } \widetilde{p}_H^{(r)} = \arg\max_{p_H} \mathcal{F}(\widetilde{p}_A^{(r-1)} \ p_H \ \widetilde{p}_Q^{(r-1)}, \widetilde{p}_Z^{(r-1)}, \boldsymbol{\Theta}^{(r-1)})$$

$$\textbf{VE-A: } \widetilde{p}_A^{(r)} = \arg\max_{p_A} \mathcal{F}(p_A \ \widetilde{p}_H^{(r)} \ \widetilde{p}_Q^{(r-1)} \ \widetilde{p}_Z^{(r-1)}, \boldsymbol{\Theta}^{(r-1)})$$

$$\textbf{VE-Q: } \widetilde{p}_Q^{(r)} = \arg\max_{p_Q} \mathcal{F}(\widetilde{p}_A^{(r)} \ \widetilde{p}_H^{(r)} \ p_Q \ \widetilde{p}_Z^{(r-1)}, \boldsymbol{\Theta}^{(r-1)})$$

$$\textbf{VE-Z: } \widetilde{p}_Z^{(r)} = \arg\max_{p_Z} \mathcal{F}(\widetilde{p}_A^{(r)} \ \widetilde{p}_H^{(r)} \ \widetilde{p}_Q^{(r)} \ p_Z, \boldsymbol{\Theta}^{(r-1)}).$$

In an equivalent way, these updates can be written as:

$$\textbf{VE-H: } \widetilde{p}_H^{(r)}(\boldsymbol{H}) \propto \exp\Big(\mathrm{E}_{\widetilde{p}_A^{(r-1)} \widetilde{p}_Z^{(r-1)}}\big[\log p(\boldsymbol{H} \,|\, \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{Z}; \boldsymbol{\Theta}^{(r-1)}\big]\Big) \tag{8}$$

$$\textbf{VE-A: } \widetilde{p}_A^{(r)}(\boldsymbol{A}) \propto \exp\Big(\mathrm{E}_{\widetilde{p}_H^{(r)} \widetilde{p}_Q^{(r-1)}}\big[\log p(\boldsymbol{A} \,|\, \boldsymbol{Y}, \boldsymbol{H}, \boldsymbol{Q}; \boldsymbol{\Theta}^{(r-1)})\big]\Big) \tag{9}$$

$$\textbf{VE-Q: } \widetilde{p}_Q^{(r)}(\boldsymbol{Q}) \propto \exp\Big(\mathrm{E}_{\widetilde{p}_A^{(r)}}\big[\log p(\boldsymbol{Q} \,|\, \boldsymbol{Y}, \boldsymbol{A}; \boldsymbol{\Theta}^{(r-1)})\big]\Big) \tag{10}$$

$$\textbf{VE-Z: } \widetilde{p}_Z^{(r)}(\boldsymbol{Z}) \propto \exp\Big(\mathrm{E}_{\widetilde{p}_H^{(r)}}\big[\log p(\boldsymbol{Z} \,|\, \boldsymbol{Y}, \boldsymbol{H}; \boldsymbol{\Theta}^{(r-1)})\big]\Big) \ . \tag{11}$$

Since $\boldsymbol{\Theta}$ and $\mathcal{G}(\widetilde{p}_{A,H,Q,Z}^{(r)})$ are independent, it follows from (7) that the **VM-step** writes:

$$\boldsymbol{\Theta}^{(r)} = \arg\max_{\boldsymbol{\Theta} \in \underline{\boldsymbol{\Theta}}} \ \mathrm{E}_{\widetilde{p}_A^{(r)} \widetilde{p}_H^{(r)} \widetilde{p}_Q^{(r)} \widetilde{p}_Z^{(r)}}\big[\log p(\boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{H}, \boldsymbol{Q}, \boldsymbol{Z}; \boldsymbol{\Theta})\big] \tag{12}$$

Compared to [21], this implies adding an E-sub-step for the HRF group assignments ($\widetilde{p}_Z^{(r)}$ updating) and specifying its impact on the other E-sub-steps. The **E-Q** sub-step ($\widetilde{p}_Q^{(r)}$ updating) is not actually impacted by the HRF groups addition and can be found in [21]. The **E-A** sub-step ($\widetilde{p}_A^{(r)}$ updating) is also very close to the one involved in [21]: similar updating formulas are obtained by replacing the HRF of [21] by voxel dependent HRFs.

It follows from standard algebra that at all iterations (r), $\widetilde{p}_H^{(r)}$ and $\widetilde{p}_A^{(r)}$ are both Gaussian distributions: $\widetilde{p}_H^{(r)} = \prod_{j \in \boldsymbol{\mathcal{V}}} \widetilde{p}_{h_j}^{(r)}$ and $\widetilde{p}_A^{(r)} = \prod_{j \in \boldsymbol{\mathcal{V}}} \widetilde{p}_{a_j}^{(r)}$, where $\widetilde{p}_{h_j}^{(r)} \sim \mathcal{N}(\boldsymbol{m}_{h_j}^{(r)}, \boldsymbol{\Sigma}_{h_j}^{(r)})$ and $\widetilde{p}_{a_j}^{(r)} \sim \mathcal{N}(\boldsymbol{m}_{a_j}^{(r)}, \boldsymbol{\Sigma}_{a_j}^{(r)})$. More specifically, the VE-steps write as follow:

- **VE-H** step: Developing the right hand side of (8), it comes that the Gaussian $\widetilde{p}_{h_j}^{(r)}$ is defined by its parameters $\boldsymbol{\Sigma}_{h_j}^{(r)} = (\boldsymbol{V}_1 + \boldsymbol{V}_2)^{-1}$ and $\boldsymbol{m}_{h_j}^{(r)} = \boldsymbol{\Sigma}_{h_j}^{(r)}(\boldsymbol{m}_1 + \boldsymbol{m}_2)$, where $\boldsymbol{V}_1 = \sum_{m,m'} \boldsymbol{\Sigma}_{a_j(m,m')}^{(r-1)} \boldsymbol{X}_m^{\mathrm{t}} \boldsymbol{\Gamma}_j^{(r-1)} \boldsymbol{X}_{m'} + \widetilde{\boldsymbol{S}}_j^{\mathrm{t}} \boldsymbol{\Gamma}_j^{(r-1)} \widetilde{\boldsymbol{S}}_j$, $\boldsymbol{V}_2 = \sum_{k=1}^{K} \widetilde{p}_{z_j}(k)^{(r-1)} \bar{\boldsymbol{\Sigma}}_k^{(r-1)-1}$, $\boldsymbol{m}_1 = \widetilde{\boldsymbol{S}}_j^{\mathrm{t}} \boldsymbol{\Gamma}_j^{(r-1)} (\boldsymbol{y}_j - \boldsymbol{P}\boldsymbol{\ell}_j^{(r-1)})$ and $\boldsymbol{m}_2 = \sum_{k=1}^{K} \bar{\boldsymbol{\Sigma}}_k^{(r-1)-1} \widetilde{p}_{z_j}(k)^{(r-1)} \bar{\boldsymbol{h}}_k^{(r-1)}$. Above, $\widetilde{\boldsymbol{S}}_j = \sum_{m=1}^{M} m_{a_j^m}^{(r-1)} \boldsymbol{X}_m$ and $m_{a_j^m}^{(r-1)}$, $\boldsymbol{\Sigma}_{a_j(m,m')}^{(r-1)}$ denote respectively the m and (m, m') entries of $\boldsymbol{m}_{a_j}^{(r-1)}$ and $\boldsymbol{\Sigma}_{a_j}^{(r-1)}$.
- **VE-A** step : This step is similar to the one of the JDE model developed in [21] if we consider a different HRF per voxel j. Using Eq. (9), standard algebra rules allow to identify the Gaussian distribution $\widetilde{p}_{a_j}^{(r)} \sim \mathcal{N}(\boldsymbol{m}_{a_j}^{(r)}, \boldsymbol{\Sigma}_{a_j}^{(r)})$, with

$$\boldsymbol{\Sigma}_{a_j}^{(r)} = \Big(\sum_{i \in \{0,1\}} \boldsymbol{\Delta}_{ij} + \widetilde{\boldsymbol{H}}_j\Big)^{-1} \tag{13}$$

$$\boldsymbol{m}_{a_j}^{(r)} = \boldsymbol{\Sigma}_{a_j}^{(r)} \Big(\sum_{i \in \{0,1\}} \boldsymbol{\Delta}_{ij} \boldsymbol{\mu}_i^{(r-1)} + \widetilde{\boldsymbol{G}}^{\mathrm{t}} \boldsymbol{\Gamma}_j^{(r-1)} \big(\boldsymbol{y}_j - \boldsymbol{P}\boldsymbol{\ell}_j^{(r-1)}\big)\Big),$$

where a number of intermediate quantities need to be specified. First, $\boldsymbol{\mu}_i^{(r-1)} = \left[\mu_{1i}^{(r-1)}, \ldots, \mu_{Mi}^{(r-1)}\right]^{\mathrm{t}}$ and $\widetilde{\boldsymbol{G}} = \mathrm{E}_{\widetilde{p}_H^{(r)}}\left[\boldsymbol{G}\right]$ where $\boldsymbol{G}$ is the matrix $\boldsymbol{G} = [\boldsymbol{g}_1 \,|\, \ldots \,|\, \boldsymbol{g}_M]$ made of columns $\boldsymbol{g}_m = \boldsymbol{X}_m \boldsymbol{h}_j$. The m-th column of $\widetilde{\boldsymbol{G}}$ is then also denoted by $\widetilde{\boldsymbol{g}}_m = \boldsymbol{X}_m \boldsymbol{m}_{h_j}^{(r)} \in \mathbb{R}^N$. Then, $\boldsymbol{\Delta}_{ij} = \mathrm{diag}_M\left[\widetilde{p}_{q_j^m}^{(r-1)}(i)/v_{mi}^{(r-1)}\right]$ and $\widetilde{\boldsymbol{H}}_j = \mathrm{E}_{\widetilde{p}_{h_j}^{(r)}}\left[\boldsymbol{G}^{\mathrm{t}}\boldsymbol{\Gamma}_j^{(r-1)}\boldsymbol{G}\right]$ is an $M \times M$ matrix whose element (m, m') is given by:

$$\begin{aligned}
&\mathrm{E}_{\widetilde{p}_{h_j}^{(r)}}\left[\boldsymbol{g}_m^{\mathrm{t}}\boldsymbol{\Gamma}_j^{(r-1)}\boldsymbol{g}_{m'}\right] = \mathrm{E}_{\widetilde{p}_{h_j}^{(r)}}\left[\boldsymbol{g}_m\right]^{\mathrm{t}}\boldsymbol{\Gamma}_j^{(r-1)}\mathrm{E}_{\widetilde{p}_{h_j}^{(r)}}\left[\boldsymbol{g}_{m'}\right] \\
&+ \mathrm{trace}\big(\boldsymbol{\Gamma}_j^{(r-1)} cov_{\widetilde{p}_{h_j}^{(r)}}(\boldsymbol{g}_m, \boldsymbol{g}_{m'})\big) \\
&= \widetilde{\boldsymbol{g}}_m^{\mathrm{t}}\boldsymbol{\Gamma}_j^{(r-1)}\widetilde{\boldsymbol{g}}_{m'} + \mathrm{trace}\big(\boldsymbol{\Gamma}_j^{(r-1)}\boldsymbol{X}_m\boldsymbol{\Sigma}_{h_j}^{(r)}\boldsymbol{X}_{m'}^{\mathrm{t}}\big).
\end{aligned}$$

- **VE-Q** step: This step remains conceptually the same as for the JDE model developed in [21] where no parcellation is estimated online. However, we provide below a slightly different and simpler presentation. In contrast to [21], we assume from the start that for all $m = 1 : M$, $\widetilde{p}_{\boldsymbol{q}^m}^{(r)}(\boldsymbol{q}^m) = \prod_{j \in \boldsymbol{\mathcal{V}}} \widetilde{p}_{q_j^m}^{(r)}(q_j^m)$ so that the VE-Q step (10) actually divides into $M \times J$ successive sub-steps. More specifically, for $m = 1 : M$, each VE-$\boldsymbol{q}^m$ step is decomposed into J successive VE-q_j^m step. It comes, for $j \in \boldsymbol{\mathcal{V}}$:

$$\begin{aligned}
&\widetilde{p}_{q_j^m}^{(r)}(q_j^m) \propto \qquad (14)\\
&\exp(\mathrm{E}_{\widetilde{p}_A^{(r)}\widetilde{p}_{q^{\setminus m}}^{(r-1)}\widetilde{p}_{q_{\setminus j}^m}^{(r-1)}}\left[\log p(q_j^m|\boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{q}_{\setminus j}^m, \boldsymbol{q}^{\setminus m}; \boldsymbol{\Theta}^{(r-1)}\right])
\end{aligned}$$

where $\boldsymbol{q}_{\setminus j}^m = \{q_{j'}^m, j' \neq j\}$ and $\boldsymbol{q}^{\setminus m} = \{\boldsymbol{q}^{m'}, m' \neq m\}$. Keeping only the terms that involve q_j^m, it follows,

$$\begin{aligned}
&\widetilde{p}_{q_j^m}^{(r)}(q_j^m) \propto \exp(\mathrm{E}_{\widetilde{p}_{a_j^m}^{(r)}}\left[\log p(a_j^m|q_j^m; \mu_m^{(r-1)}, v_m^{(r-1)}\right] \\
&+ \mathrm{E}_{\widetilde{p}_{\boldsymbol{q}_{\setminus j}^m}^{(r-1)}}\left[\log p(q_j^m|\boldsymbol{q}_{\setminus j}^m; \beta_m^{(r-1)}\right]) \qquad (15)\\
&\propto \exp\left(-\frac{1}{2}\frac{(\boldsymbol{m}_{a_j^m}^{(r)} - \mu_{mq_j^m}^{(r-1)})^2 + \boldsymbol{\Sigma}_{a_j(m,m)}^{(r)}}{v_{mq_j^m}^{(r-1)}}\right) \\
&\times \exp\left(-\frac{1}{2}\log v_{mq_j^m}^{(r-1)} + \beta_m^{(r-1)} \sum_{j' \in \boldsymbol{\mathcal{V}}, j' \sim j} \widetilde{p}_{q_{j'}^m}^{(r-1)}(q_j^m)\right) \qquad (16)
\end{aligned}$$

For $q_j^m = i$, it follows

$$\begin{aligned}
&\widetilde{p}_{q_j^m}^{(r)}(i) \propto \mathcal{N}(\boldsymbol{m}_{a_j^m}^{(r)}; \mu_{mi}^{(r-1)}, v_{mi}^{(r-1)}) \times \qquad (17)\\
&\exp\left(-\frac{1}{2}\frac{\boldsymbol{\Sigma}_{a_j(m,m)}^{(r)}}{v_{mi}^{(r-1)}} + \beta_m^{(r-1)} \sum_{j' \in \boldsymbol{\mathcal{V}}, j' \sim j} \widetilde{p}_{q_{j'}^m}^{(r-1)}(i)\right),
\end{aligned}$$

where the Gaussian density function with mean μ_{mi} and variance v_{mi} is denoted by $\mathcal{N}(\,.\,; \mu_{mi}, v_{mi})$. Note that the above expression corresponds to a synchronous updating of each $\widetilde{p}_{q_j^m}$ in parallele from the previous $\widetilde{p}_{q_j^m}^{(r-1)}$'s. However in practice, as soon as $\widetilde{p}_{q_j^m}$ is updated for some j, it will be taken into account in the following updates of the subsequent $\widetilde{p}_{q_{j'}^m}$ for $j' \neq j$.

- **VE-Z** step: As for the VE-Q step, we look for a variational solution of the form $\widetilde{p}_Z^{(r)}(\boldsymbol{Z}) = \prod_{j \in \boldsymbol{\mathcal{V}}} \widetilde{p}_{z_j}^{(r)}(z_j)$, which leads to J successive sub-steps performed in turn. For each $j \in \boldsymbol{\mathcal{V}}$, and $k = 1 : K$,

$$\begin{aligned}
&\widetilde{p}_{z_j}^{(r)}(k) \propto \mathcal{N}(\boldsymbol{m}_{h_j}^{(r)}; \bar{\boldsymbol{h}}_k^{(r-1)}, \bar{\boldsymbol{\Sigma}}_k^{(r-1)}) \times \qquad (18)\\
&\exp\left(-\frac{1}{2}\mathrm{trace}(\boldsymbol{\Sigma}_{h_j}^{(r)}\bar{\boldsymbol{\Sigma}}_k^{(r-1)-1}) + \beta_z^{(r-1)} \sum_{j' \in \boldsymbol{\mathcal{V}}, j' \sim j} \widetilde{p}_{z_{j'}}^{(r-1)}(k)\right).
\end{aligned}$$

- **VM** step: For this step, we can first rewrite Eq. (12) as

$$\begin{aligned}
&\boldsymbol{\Theta}^{(r)} = \arg\max_{\boldsymbol{\Theta} \in \underline{\boldsymbol{\Theta}}} \left[\mathrm{E}_{\widetilde{p}_A^{(r)}\widetilde{p}_H^{(r)}}\left[\log p(\boldsymbol{Y} \,|\, \boldsymbol{A}, \boldsymbol{H}; \boldsymbol{L}, \boldsymbol{\Gamma})\right] + \right. \\
&\mathrm{E}_{\widetilde{p}_A^{(r)}\widetilde{p}_Q^{(r)}}\left[\log p(\boldsymbol{A} \,|\, \boldsymbol{Q}; \boldsymbol{\mu}, \boldsymbol{v})\right] + \mathrm{E}_{\widetilde{p}_Q^{(r)}}\left[\log p(\boldsymbol{Q}; \boldsymbol{\beta})\right] + \\
&\mathrm{E}_{\widetilde{p}_Z^{(r)}}\left[\log p(\boldsymbol{Z}; \beta_z)\right] + \mathrm{E}_{\widetilde{p}_H^{(r)}\widetilde{p}_Z^{(r)}}\left[\log p(\boldsymbol{H} \,|\, \boldsymbol{Z}; \boldsymbol{\nu}, \bar{\boldsymbol{h}})\right] + \\
&\left.\log p(\bar{\boldsymbol{h}}; \sigma_h^2)\right]. \qquad (19)
\end{aligned}$$

The maximization step can therefore be divided into five sub-steps (two additional ones compared to [21]) involving separately $(\boldsymbol{\mu}, \boldsymbol{v})$, $\boldsymbol{\beta}$, β_z, $(\mathbf{L}, \boldsymbol{\Gamma})$, σ_h^2 and $(\boldsymbol{\nu}, \bar{\boldsymbol{h}})$.

- **VM-$(\boldsymbol{\mu}, \boldsymbol{v})$** step: This step is equivalent to the one in [21, Appendix E-1)].
- **VM-$(\boldsymbol{L}, \boldsymbol{\Gamma})$** step: This step is also equivalent to the one in [21, Appendix E-4)].
- **VM-$\boldsymbol{\beta}$** step: By maximizing with respect to $\boldsymbol{\beta}$, Eq. (19) reads:

$$\boldsymbol{\beta}^{(r)} = \arg\max_{\boldsymbol{\beta}} \mathrm{E}_{\widetilde{p}_Q^{(r)}}\left[\log p(\boldsymbol{Q}; \boldsymbol{\beta})\right]. \qquad (20)$$

Using additional exponential priors (with respective parameter λ_{β_m}) akin to [21] in order to avoid over-estimation of this parameter and hence having extra spatial regularization, this M-step can be reformulated as, for each $m = 1 : M$,

$$\beta_m^{(r)} = \arg\max_{\beta_m} E_{\widetilde{p}_{q^m}^{(r)}}\left[\log p(\boldsymbol{q}^m; \beta_m)\right] + \log p(\beta_m; \lambda_{\beta_m}). \qquad (21)$$

The optimal value of $\boldsymbol{\beta}^{(r)}$ can be reached following an iterative procedure as mentioned in [21]. We provide here the specific implementation used in this paper. This step does not admit an explicit closed-form expression but can be solved numerically using gradient ascent schemes. It is straightforward to show that the maximization of (21) in β_m is equivalent to solve

$$\beta_m^{(r)} = \arg\max_{\beta_m} \beta_m(E_{\widetilde{p}_{q^m}^{(r)}}\left[U(\boldsymbol{q}^m)\right] - \lambda_{\beta_m}) - \log \mathcal{K}(\beta_m) \qquad (22)$$

which admits a unique solution, and where $\mathcal{K}$ denotes the normalizing constant that depends on β_m. The first and second derivatives with regards to β_m can be easily computed as respectively:

$$\nabla_{\beta_m} = E_{\widetilde{p}_{q^m}^{(r)}}\left[U(\boldsymbol{q}^m)\right] - E_{p(q^m;\beta_m)}\left[U(\boldsymbol{q}^m)\right] - \lambda_{\beta_m} \qquad (23)$$

$$\nabla_{\beta_m}^2 = -var_{p(q^m;\beta_m)}\left[U(\boldsymbol{q}^m)\right] \leq 0 \qquad (24)$$

where $p(q^m; \beta_m)$ is the Potts model defined in (3) and $var_{p(q^m;\beta_m)}$ denotes the variance with respect to (w.r.t.) $p(q^m; \beta_m)$. The function to optimize is thus concave. Unfortunately, due to the intractable normalizing constant $\mathcal{K}$, the two expressions above are not directly available. It is then necessary to approximate the terms involving the true MRF prior $p(\boldsymbol{q}^m; \beta_m)$. A simple solution is to use a Mean Field like approximation as presented in [29] in which the spatial interactions between neighboring voxels are made

tractable by replacing the stochastic neighbors with fixed values:

$$p^{MF}_{q^m}(\boldsymbol{q}^m;\beta_m)=\prod_{j\in\boldsymbol{\mathcal{V}}}p^{MF}_{q^m_j}(q^m_j;\beta_m) \tag{25}$$

with $p^{MF}_{q^m_j}(q^m_j;\beta_m)$ defined by:

$$p^{MF}_{q^m_j}(q^m_j=i;\beta_m)=\frac{\exp(\beta_m\sum\limits_{j'\in\boldsymbol{\mathcal{V}},j'\sim j}\widetilde{p}^{(r)}_{q^m_{j'}}(i))}{\sum\limits_{l\in\{0,1\}}\exp(\beta_m\sum\limits_{j'\in\boldsymbol{\mathcal{V}},j'\sim j}\widetilde{p}^{(r)}_{q^m_{j'}}(l))} \tag{26}$$

It follows,

$$\begin{aligned}\nabla_{\beta_m}&\approx\sum_{(j,j')\in\boldsymbol{\mathcal{V}}^2,j\sim j'}\sum_{i\in\{0,1\}}(\widetilde{p}^{(r)}_{q^m_j}(i)\widetilde{p}^{(r)}_{q^m_{j'}}(i) \\ &\quad-p^{MF}_{q^m_j}(i;\beta_m)p^{MF}_{q^m_{j'}}(i;\beta_m))-\lambda_{\beta_m} \\ &\approx\frac{1}{2}\sum_{i\in\{0,1\}}\sum_{j\in\boldsymbol{\mathcal{V}}}\Bigg(\widetilde{p}^{(r)}_{q^m_j}(i)\sum_{j'\in\boldsymbol{\mathcal{V}},j'\sim j}\widetilde{p}^{(r)}_{q^m_{j'}}(i) \\ &\quad-p^{MF}_{q^m_j}(i;\beta_m)\sum_{j'\in\boldsymbol{\mathcal{V}},j'\sim j}p^{MF}_{q^m_{j'}}(i;\beta_m)\Bigg)-\lambda_{\beta_m}\end{aligned} \tag{27}$$

where we emphasize that the dependence in β^m is in the second term, the other terms being constant. In both $\widetilde{p}^{(r)}_{q^m_j}(i)$ and $p^{MF}_{q^m_j}(i;\beta_m)$, the Markov part is treated similarly using the *best* currently available posterior variational approximation $\widetilde{p}^{(r-1)}_{q^m}$ or $\widetilde{p}^{(r)}_{q^m}$. The main difference between $\widetilde{p}^{(r)}_{q^m_j}(i)$ and $p^{MF}_{q^m_j}(i;\beta_m)$ is that $p^{MF}_{q^m_j}(i;\beta_m)$ does not involve directly terms that depend on the observed data $\boldsymbol{Y}$ which is consistent with the fact that it represents an approximation of the prior MRF.

- **VM-**β_z step: By maximizing with respect to β_z, Eq. (19) reads:

$$\beta_z^{(r)}=\arg\max_{\beta_z}\ \mathrm{E}_{\widetilde{p}^{(r)}_Z}\big[\log p(\boldsymbol{Z};\beta_z)\big]. \tag{28}$$

To penalize high values of β_z to avoid over spatial regularization, an exponential prior can also be used as for the M-$\boldsymbol{\beta}$ step. The above optimization procedure can therefore be reformulated as

$$\beta_z^{(r)}=\arg\max_{\beta_z}E_{\widetilde{p}^{(r)}_Z}[\log p(\boldsymbol{Z};\beta_z)]+\log p(\beta_z;\lambda_z) \tag{29}$$

where λ_z is the parameter of the exponential distribution. The same iterative procedure as for the M-$\boldsymbol{\beta}$ step can be used to reach the optimal value of β_z making use of the possibility to approximate the gradient as

$$\begin{aligned}\nabla_{\beta_z}&\approx\sum_{(j,j')\in\boldsymbol{\mathcal{V}}^2,j\sim j'}\sum_{k=1:K}(\widetilde{p}^{(r)}_{z_j}(k)\widetilde{p}^{(r)}_{z_{j'}}(k) \\ &\quad-p^{MF}_{z_j}(k;\beta_z)p^{MF}_{z_{j'}}(k;\beta_z))-\lambda_{\beta_z},\end{aligned} \tag{30}$$

with

$$p^{MF}_{z_j}(z_j=k;\beta_z)=\frac{\exp(\beta_z\sum\limits_{j'\in\boldsymbol{\mathcal{V}},j'\sim j}\widetilde{p}^{(r)}_{z_{j'}}(k))}{\sum\limits_{l=1:K}\exp(\beta_z\sum\limits_{j'\in\boldsymbol{\mathcal{V}},j'\sim j}\widetilde{p}^{(r)}_{z_{j'}}(l))} \tag{31}$$

- **VM-**$(\sigma_h^2,\bar{\boldsymbol{h}},\boldsymbol{\nu})$ step: This maximization step can be reached by solving the following problem

$$\begin{aligned}\left(\sigma_h^{2\,(r)},\bar{\boldsymbol{h}}^{(r)},\boldsymbol{\nu}^{(r)}\right)&=\arg\max_{\sigma_h^2,\bar{\boldsymbol{h}},\boldsymbol{\nu}}\mathrm{E}_{\widetilde{p}^{(r)}_H\widetilde{p}^{(r)}_Z}\big[\log p(\boldsymbol{H}\mid\boldsymbol{Z};\bar{\boldsymbol{h}},\boldsymbol{\nu})\big] \\ &\quad+\log p(\bar{\boldsymbol{h}};\sigma_h^2).\end{aligned} \tag{32}$$

Derivations with respect to the $\bar{\boldsymbol{h}}_k$'s and $\boldsymbol{\nu}_k$'s provide the following equations for a given k

$$\boldsymbol{\nu}_k=\frac{\sum\limits_{j\in\boldsymbol{\mathcal{V}}}\widetilde{p}^{(r)}_{z_j}(k)\left(\mathrm{trace}(\boldsymbol{\Sigma}^{(r)}_{h_j})+(\boldsymbol{m}^{(r)}_{h_j}-\bar{\boldsymbol{h}}_k)^{\mathrm{t}}(\boldsymbol{m}_{h_j}-\bar{\boldsymbol{h}}_k)\right)}{(D+1)\sum\limits_{j\in\boldsymbol{\mathcal{V}}}\widetilde{p}^{(r)}_{z_j}(k)} \tag{33}$$

and

$$\bar{\boldsymbol{h}}_k=\left(I_{D+1}+\frac{\boldsymbol{\nu}_k\boldsymbol{R}^{-1}/\sigma_h^2}{\sum\limits_{j\in\boldsymbol{\mathcal{V}}}\widetilde{p}^{(r)}_{z_j}(k)}\right)^{-1}\frac{\sum\limits_{j\in\boldsymbol{\mathcal{V}}}\widetilde{p}^{(r)}_{z_j}(k)\boldsymbol{m}^{(r)}_{h_j}}{\sum\limits_{j\in\boldsymbol{\mathcal{V}}}\widetilde{p}^{(r)}_{z_j}(k)} \tag{34}$$

Then updated values $\boldsymbol{\nu}_k^{(r)}$ and $\bar{\boldsymbol{h}}_k^{(r)}$ can be obtained by solving equations (33) and (34). An estimation of σ_h^2 follows as

$$\sigma_h^{2(r)}=\frac{\sum_{k=1}^K\bar{\boldsymbol{h}}_k^{(r)t}\boldsymbol{R}^{-1}\bar{\boldsymbol{h}}_k^{(r)}}{K} \tag{35}$$

However, as σ_h^2 plays the role of a smoothing parameter for the HRF shapes, it will be held fixed in the present work. Note that in the more general case of non isotropic $\bar{\boldsymbol{\Sigma}}_k$, we would get

$$\bar{\boldsymbol{\Sigma}}_k=\frac{\sum\limits_{j\in\boldsymbol{\mathcal{V}}}\widetilde{p}^{(r)}_{z_j}(k)\left(\boldsymbol{\Sigma}^{(r)}_{h_j}+(\boldsymbol{m}^{(r)}_{h_j}-\bar{\boldsymbol{h}}_k)(\boldsymbol{m}^{(r)}_{h_j}-\bar{\boldsymbol{h}}_k)^{\mathrm{t}}\right)}{\sum\limits_{j\in\boldsymbol{\mathcal{V}}}\widetilde{p}^{(r)}_{z_j}(k)} \tag{36}$$

and

$$\bar{\boldsymbol{h}}_k=\left(I_{D+1}+\frac{\bar{\boldsymbol{\Sigma}}_k\boldsymbol{R}^{-1}/\sigma_h^2}{\sum\limits_{j\in\boldsymbol{\mathcal{V}}}\widetilde{p}^{(r)}_{z_j}(k)}\right)^{-1}\frac{\sum\limits_{j\in\boldsymbol{\mathcal{V}}}\widetilde{p}^{(r)}_{z_j}(k)\boldsymbol{m}_{h_j}}{\sum\limits_{j\in\boldsymbol{\mathcal{V}}}\widetilde{p}^{(r)}_{z_j}(k)}. \tag{37}$$

The first case (33) can be obtained from the general one (36) by setting $\boldsymbol{\nu}_k=\mathrm{trace}(\bar{\boldsymbol{\Sigma}}_k)/(D+1)$.

IV. Validation of the proposed approach

A. Validation on synthetic data

1) Case of study: Experiments have been carried out on artificial fMRI data generated according to Eq. (1). We simulated a random mixed sequence of indexes coding for $M=2$ different stimuli composed of 30 trials each with a time of repetition set to $TR=1s$. The resulting sequence was then multiplied by stimulus-dependent and space-varying NRLs, which were drawn from the prior distribution $p(\boldsymbol{A};\boldsymbol{\theta}_a)$. To this end, 2D slices composed of 20×20 binary labels $\boldsymbol{q}^m$ (activating and non-activating voxels) were constructed for each stimulus type m (see Fig. 2[top]). Given these labels, the NRLs were simulated as follows, for $m=1,2$: $(a^m_j\,|\,q^m_j=0)\sim\mathcal{N}(0,0.5)$ and $(a^m_j\,|\,q^m_j=1)\sim\mathcal{N}(3.2,0.5)$ (see Fig. 5[top]). As regards HRFs, three groups ($K=3$) were considered and spatially organized in three parcels of similar size (labels $\boldsymbol{Z}$) as shown in Fig. 3. Within each parcel, all voxels share the same HRF prior parameters $(\bar{\boldsymbol{h}}_k,\bar{\boldsymbol{\Sigma}}_k)$. The mean HRF shapes $\{\bar{\boldsymbol{h}}_k,k=1:K\}$ are depicted in Fig. 4 (ground truth) and show strong fluctuations across parcels in terms of peak positions and widths. Isotropic prior covariance matrices $\{\bar{\boldsymbol{\Sigma}}_k=\nu_kI_{D+1},k=1:K\}$ were considered to draw voxel-specific HRFs according to $p(\boldsymbol{h}_j|z_j=k)$ where ν_k has been set to 0.02.

As regards parcellation, Fig. 3[center] shows the ability of JPDE to recover the spatial support of hemodynamic territories with high accuracy (1% of misclassified voxels and a DICE index [30] of

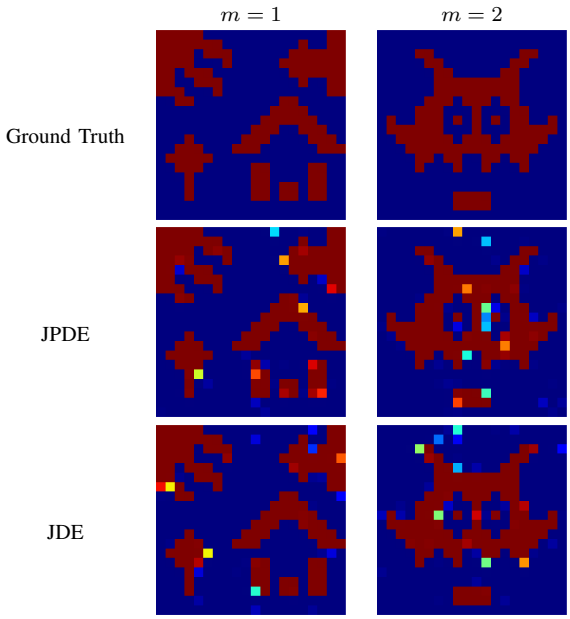

Fig. 2. Reference activation labels and Posterior Probability Maps (PPM) for JPDE and JDE (a single parcel is assumed for JDE).

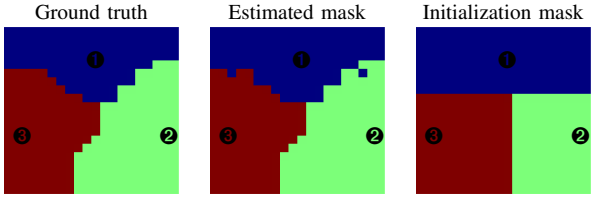

Fig. 3. Ground truth (left), JPDE estimated (center) and the initial parcellation (right) masks.

0.993) from an arbitrary initialization (Fig. 3[right]). The HRF variability does not seem to affect the activation maps which are equally well estimated in the JPDE and JDE cases (Fig. 2[center] and Fig. 2[bottom]). However, a clear difference is seen on the estimated HRFs, which are depicted in Fig. 4 together with the ground truth: the three parcel-specific HRF estimates using JPDE are plotted in addition to the JDE-based HRF time course obtained by merging all parcels in the arbitrary initial parcellation. This logically provides a single HRF time course for JDE inference. Then, three complementary JDE experiments have been conducted, where in turn each of the three ground truth HRFs depicted in Fig. 4 has been used as global reference HRF in a separate experiment. The HRF estimates are then compared to the ground truth and to the JPDE solution associated with the corresponding parcel. We observed that the JPDE estimation is accurate in all parcels although the parcels cover different proportions of activation areas (*i.e.* the amount of useful signal varies from one parcel to another). In contrast, JDE provides an intermediate HRF shape that lies between those of the three parcels. This explains the observed differences between the two models in terms of NRL estimates and points out the JDE sensitivity to the choice of the a priori parcellation. When the parcellation is imperfect, JDE enforces fitting a wrong HRF model to fMRI time series, and therefore a wrong activation dynamics. In contrast, JPDE is able to automatically refine an initial candidate for the parcellation and provides reliable detection and estimation results.

Interestingly, the NRL differences in Fig. 5 (see the JPDE-JDE plots in Fig. 5[bottom]) show that NRL estimates with JPDE have higher values, which means that JPDE allows retrieving stronger activation dynamics closer to the ground truth.

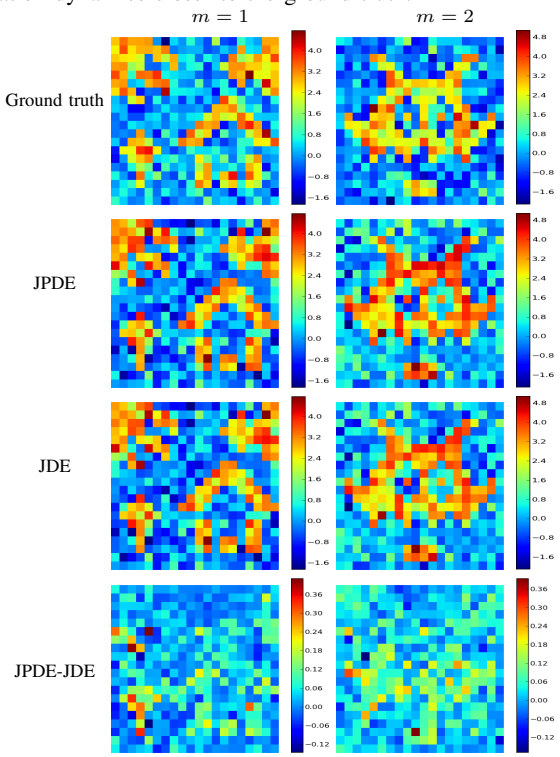

Fig. 5. Reference and estimated NRLs using JPDE (3 parcels) and JDE (1 parcel).

The most significant NRL differences lie in parcels 2 and 3 where the JDE HRF estimate differs the most from the ground truth. In terms of Mean Square Error (MSE), reported values confirm the improved performance of JPDE over JDE: $MSE^{m=1}_{\text{JDE}} = 0.0182$ *vs* $MSE^{m=1}_{\text{JPDE}} = 0.0107$ and $MSE^{m=2}_{\text{JDE}} = 0.0183$ *vs* $MSE^{m=2}_{\text{JPDE}} = 0.0141$.

2) Robustness study: The aim of this part is to assess the robustness of the JPDE model to a number of confounding factors. We first investigated the robustness of JPDE to the input SNR in terms of estimation error on the output NRLs. We then investigated the robustness of JPDE to the initial parcellation in terms of DICE coefficient for the estimated output parcellation. These first two experiments were conducted on 50 runs in order to compute mean and standard deviation values. A third set of experiments was designed to measure the impact of the overlapping percentage between the parcellation and activation maps in order to study how the presence of activations within each parcel helped to achieve good parcellation results.

a) Robustness to input SNR: In this experiment, 50 simulations have been used to evaluate the output MSE on NRL estimates for five increasing values of input SNR. Also, a 3-parcel model has been used with the same ground truth with respect to activation masks and parcellation as in Figs. 2-3, respectively. Fig. 6 shows the evolution of the mean MSE on NRL estimates with respect to the input SNR in addition to the standard deviation over this mean value. Mean MSE and standard deviations are displayed separately (color coded) for each experimental condition m. Our results indicate low error level on NRL estimates and a decreasing trend of the mean MSE as the input SNR increases. We concomitantly observed an improved stability (lower standard deviations) as the input SNR increases. These results are consistent with those observed in our previous work on the JDE model [21].

b) Robustness to initial parcellation: Here, we investigated the sensitivity of the JPDE model to the initial parcellation mask.

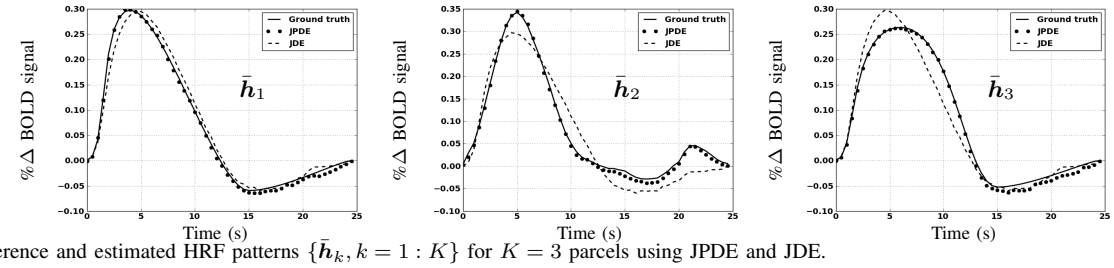

Fig. 4. Reference and estimated HRF patterns $\{\bar{h}_k, k = 1 : K\}$ for $K = 3$ parcels using JPDE and JDE.

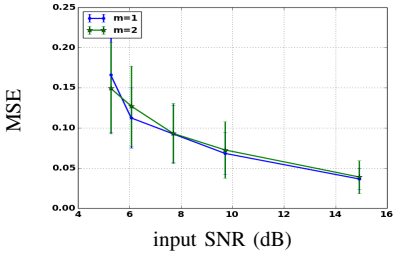

Fig. 6. Mean MSE and standard deviation on estimated NRLs with JPDE for 5 different input SNR values and for the two experimental conditions.

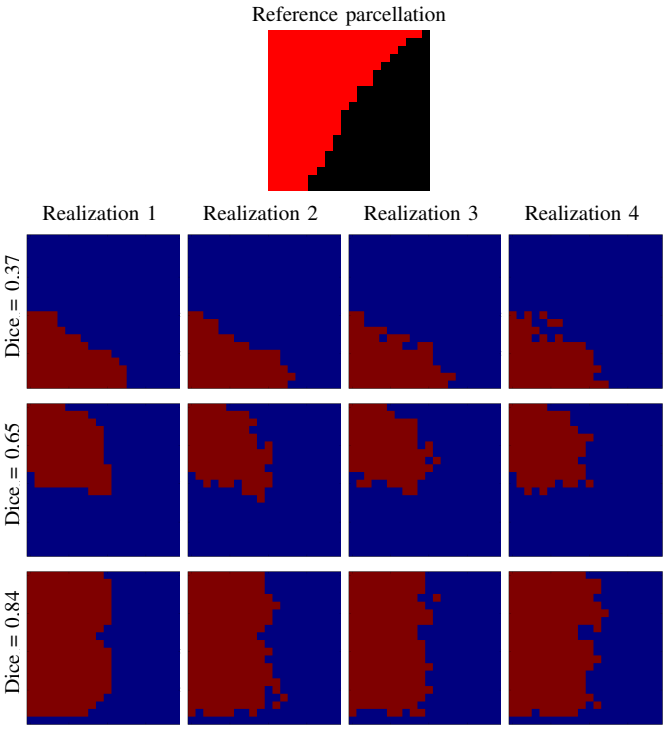

Fig. 7. Reference parcellation mask (top row) and pseudo-randomly generated parcellations with a 2-parcel model. Displayed masks for three DICE coefficient groups. Each row illustrates different realizations of masks sharing the same DICE coefficient.

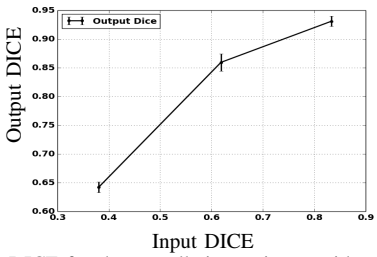

Fig. 8. Output DICE for the parcellation estimate with respect to the input DICE over 50 runs for the three initial parcellation masks shown in Fig. 7.

The accuracy of the initial parcellation mask, which is given as input to JPDE, may actually have a strong impact on the final estimated parcellation. The accuracy was measured in terms of DICE coefficients for each mask estimate. More specifically, different DICE levels between the ground truth (reference parcellation mask, see Fig. 7) and the initial parcellation have been tested to cover a large number of situations. Over 50 runs, for each DICE level, different initial parcellation masks were given to the JPDE algorithm. These masks were generated in a pseudo-random manner so as to guarantee a varying border between two parcels in a 2-parcel model, while approximately ensuring connexity and the same DICE coefficient value for any realization of such masks (see Fig. 7). In contrast to previous subsections, since here we are only considering 2 parcels, we restrict the HRF groups to the first two shown in Fig. 4 where $\bar{h}_1$ and $\bar{h}_2$ are associated with the red and black parcels shown in Fig. 7[top], respectively. More details about this pseudo-random synthesis are given in[31, Appendix B].

Fig. 8 shows the DICE coefficient (mean value and standard deviation) for the final estimated parcellation with respect to the DICE level for the initial parcellation. Through the three measured mean points, one can first notice that the output mean DICE value increases with the input DICE, which was expected since the more accurate the initialization the better the final parcellation estimate is. Second, as regards variability measured through the standard deviation, it remains almost the same. One can however notice a slightly larger variability in the second example (DICE $= 0.65$). The reason is that the red parcel in this example (third row in Fig. 8) mainly overlaps with an activated area from the first experimental condition ($m = 1$). The different realizations in this example actually involve only a few activated voxels for the second experimental condition ($m = 2$). This aspect is specifically addressed in the following part.

c) Activation coverage: Owing to the hierarchical model adopted for JPDE, parcels are estimated (parcellation task) from the HRF patterns (estimation task), which are strongly linked to the accuracy of detected activations (detection task). These three tasks are thus strongly linked to one another within the JPDE formulation. For this reason, and as already shown in previous works on the standard JDE model [19–21], accurate HRF estimation is mainly expected in activated brain areas, and vice versa, improved activation detection should be achieved from more accurate HRF estimates. In the same vein, accurate delineation of parcels is mainly expected if the latter embody evoked brain activity. The presence of activations within each target parcel is therefore an important concern that is now investigated. To this end, three numerical experiments have been conducted using different configurations for the reference parcellation. A 2-parcel model has been used with the same activation maps as in Fig. 2 and still the first two HRF groups. They are superimposed to the fusion of activation maps ($m = 1$ and $m = 2$), which has been obtained using a logical "OR" applied to the two activation

label maps. This allows us to compute the overlapping percentage between the two parcels and this resulting *global activation map*, which changes over numerical experiments (see Fig. 9).

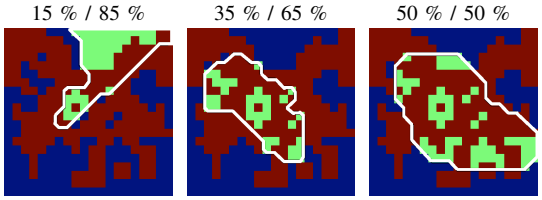

Fig. 9. Overlaps between the parcellation masks (green parcel delimited by the white line) and the *global activation map* (red). The overlapping percentages between the two parcels and this global map are reported on top of each mask.

We then focused on the parcellation results and evaluated them in terms of output DICE index. Fig. 10 provides the output DICE curve w.r.t. the overlapping percentage between the parcels and the global activation map. Slightly different masks were given as initialization to the JPDE algorithm based on the pseudo-random generation detailed in Section IV-A2b. We first observed that a fair partition (50%/50%) of activations across parcels helped to achieve better parcellation performance compared to unbalanced distribution of activations (eg. 15%/85%). This result confirms our previous observations in Subsection IV-A2b: the presence of activations helps to uncover accurate brain parcellations based on hemodynamic features.

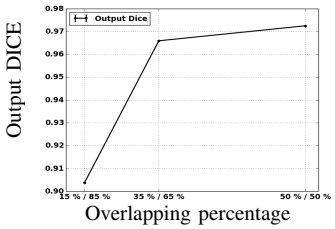

Fig. 10. Output DICE index as a function of the overlapping percentage between the global activation map and parcel definition in a 2-parcel model, where the parcellation masks are those defined in Fig. 9.

B. Validation on real data

This section is devoted to the validation of the proposed JPDE model on real fMRI data. We analyzed fMRI data acquired in a study dedicated to uncovering the large-scale organization of spoken language areas in dyslexic children [32].

1) Experimental paradigm: The paradigm consisted of a slow event-related design comprising 40 short sentences in French (native language) and Japanese (a foreign language that none of the children understood), randomly presented every 12 s. Each sentence was repeated once in a row to study repetition effect [33] and the sentences were produced by different native woman speakers. The mean sentence duration was similar in both languages (2707 ms vs 2724 ms). The paradigm comprised 4 runs, each run consisting of 4 different trials in each language. A trial was composed of two consecutive instances of the same sentence. Hence, the total duration of each run was 3 min 12 s during which 16 sentences were delivered to children. In this period of time, $N = 80$ volumes were acquired. MRI structural (TR = 2300 ms, TE = 4.18 ms, matrix 256×256×176, voxel size = 1×1×1 mm) and functional (TR = 2400 ms, TE = 30 ms, matrix, 64×64×40, voxel size =3×3×3mm) whole brain data were acquired on a 3.0 Tesla Siemens Tim Trio scanner.

2) Results: For each run separately ($N = 80$ scans), we fitted the JDE (VEM version [21]) and JPDE models using $M = 4$ experimental conditions corresponding to the first and second presentations of each sentence, either spoken in French or in Japanese. We performed such analysis in the STS mask of the left hemisphere (see Fig. 11[center]). Next, we averaged the results over the 4 runs. Hereafter, we report the JDE and JPDE averaged results for a single six-year old child. Similar findings have been observed on other individuals. This allows us to investigate whether inferring the hemodynamic parcellation performs better than considering a fixed functional segmentation. Our comparison then relies on two key properties: (i) the brain regions known to be involved in language processing and (ii) the known specificity of activation dynamics. As regards (i), six regions of interest that lie along the left STS (see Fig. 11), are emphasized according to anatomical criteria:

- A central region located in the Heschl's gyrus, and denoted by Heschl;
- A region located in the middle part of the STS, corresponding to the primary auditory cortex, referred to as middle STS and denoted by mSTS;
- Two regions in the posterior part of the STS, the most posterior being denoted by pSTS and the other one by intermediate pSTS;
- Two regions in the anterior part of the STS, the most anterior one being denoted by aSTS and the other one by intermediate aSTS.

As regards (ii), it has been shown in [33] that infants present a gradient in the dynamics of activations along the left STS, meaning that some regions respond earlier than others or that the time-to-peak (TTP) of the corresponding HRF profile is shorter[1]. This gradient has also been replicated in children [34] and adults [35]. In each case, the fastest regions were found around Heschl's gyrus, with a slow-down along the dorso-ventral and rostral-caudal axes. The more anterior and posterior parts of the left STS are known to be the slowest ones. Our results will thus be interpreted with respect to this state-of-the-art.

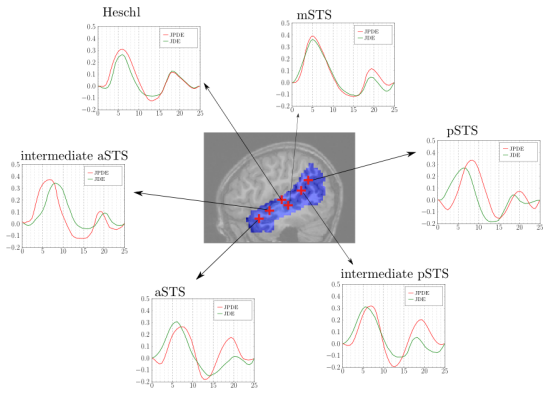

Fig. 11. The HRF profiles estimated in six regions along the left superior temporal sulcus using the JPDE (red) and JDE (green) models. In the center, the STS mask is superimposed on an anatomical slice of the 6-year old child (sagittal view, x=48). The region centers are indicated by red marks.

For JDE-based inference, we used a parcellation of the STS (see Fig. 12(a)) generated by a spatially constrained hierarchical clustering (Ward) algorithm of functional features (effects maps), the latter being extracted from a classical GLM analysis [2]. The number of parcels was set to 10. After fitting the JDE model to the fMRI data, parcel-specific HRF shapes were computed (see Fig. 11). For JPDE-based inference, the initial parcellation was set up using the same ascendant hierarchical clustering except that some noise was randomly introduced (cf. Fig. 13(a)): the number of parcels K in

[1]The TTP corresponds to the time to reach the maximum of the HRF curve.

this initial parcellation was varied from 10 to 20 corresponding to a slight over-segmentation compared to the JDE setup. Our motivation for changing K was to assess the robustness of JPDE for retrieving a coherent parcellation from this noisy initial guess and to see to what extent the corresponding HRF shapes were well recovered (Fig. 11). In what follows, we report the best results we got regarding the hemodynamic gradient, namely for $K = 13$.

First, it is worth noting that the JPDE model was able to recover a regular parcellation, as shown in Fig. 13(b). Hence, the parcellation yielded by the JPDE model not only differs from its noisy initialization (Fig. 13 (a)) but also from the fixed parcellation used in JDE (Fig. 12 (a)). Second, significant differences between JDE and JPDE models can be seen on their respective HRF profiles, whose discrepancies are reported in Fig. 11. Such differences are actually related to the parcel support changes between the two competing approaches. In Fig. 11, for the sake of clarity, we show HRF profiles for only six parcels along the STS. We observed similar time courses in the center of the STS, namely in Heschl's gyrus and middle STS, whereas we found medium to large shape differences in its anterior and posterior parts. We noticed the presence of a second HRF peak about 18 to 20 s after the first sentence onset. This peaks occurs because of the sentence repetition effect 12 s after the presentation of the first stimulus, even though we modelled as two separate conditions the first and second sentence presentations. This effect might have been mitigated either considering a slower design or introducing some jittering between the corresponding onsets. In what follows, we will restrict our TTP analysis to the study of the first peak although a specific extension of the JDE formulation has been proposed in the past to deal with repetition suppression effect [36].

Next, to assess the recovery of the temporal gradient along the STS, we compared the parcel-specific TTPs between HRF estimates delivered by the two approaches. Figs. 12(c)-13(c) report the TTP maps in a sagittal slice for the JDE and JPDE models, respectively. The colors closer to red (respectively, purple) correspond to smaller (respectively, larger) TTPs (i.e. to faster and slower responses, respectively). The faster TTP is close to 5 s whereas the slower is around 8.5 s. The JPDE model recovers a clear gradient of response, from the middle STS regions (faster responses) to the anterior and posterior parts (slower responses). In the posterior and anterior regions (pSTS and aSTS), the corresponding hemodynamic responses are significantly slower than the one estimated by JDE inference ($\text{TTP}_{\text{JPDE}} = 8.5$ s vs $\text{TTP}_{\text{JDE}} = 6.5$ s in pSTS, $\text{TTP}_{\text{JPDE}} = 7.5$ s vs $\text{TTP}_{\text{JDE}} = 6$ s in aSTS). The same observation was replicated in the intermediate posterior STS region. Furthermore, a slower HRF was retrieved in the more anterior region ($\text{TTP}_{\text{JPDE}} = 8$ s) than in the intermediate aSTS ($\text{TTP}_{\text{JPDE}} = 7$ s). Hence, in these regions, the JPDE model provides more coherent results with the existing literature. In contrast, this gradient assumption is less tenable when analyzing the JDE results. In particular, JDE-based HRF time courses revealed a TTP decrease down to 6 s in the anterior part of the STS. In the same vein, the HRF profile was slower in the intermediate aSTS ($\text{TTP}_{\text{JDE}} = 8$ s) than in the aSTS ($\text{TTP}_{\text{JDE}} = 6$ s).

Globally, using $K = 13$ the JPDE approach recovers a more coherent final parcellation of fMRI data where the HRF shapes are more accurately estimated than the ones yielded by the JDE model.

V. Discussion

Our simulated and real fMRI data experiments allowed us to investigate the performance of the JPDE framework in establishing an accurate brain parcellation directly from the data. Specifically, simulated data results demonstrated the robustness of the JPDE model with respect to a number of factors linked to the data acquisition process such as the input SNR or induced by algorithmic requirements such as the choice of an initial parcellation. We also explored how the overlap between activated areas and the different parcels impacted the algorithm performance. As regards real data, our results showed the ability of the JPDE model to recover a meaningful hemodynamic parcellation that accurately captures the temporal gradient along the STS during a speech perception experiment.

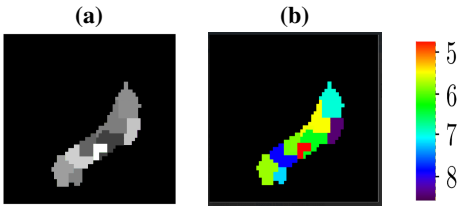

Fig. 12. Results obtained using the JDE model. **(a)**: Parcellation provided as input mask to the JDE algorithm. **(b)**: TTP map indicating the time-to-peak in s. for each parcel. The color bar encodes the shortest and longest TTPs in s. in red and purple, respectively.

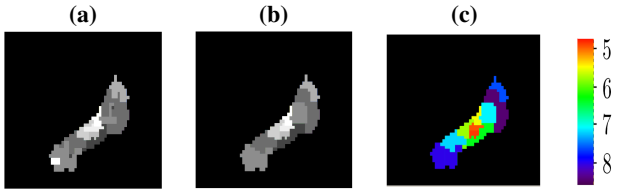

Fig. 13. Results obtained using the JPDE model. **(a)**: Initial parcellation provided as input. **(b)**: Final parcellation computed by the algorithm. **(c)**: TTP map indicating the time-to-peak in s. in each parcel of the final parcellation.

Several issues have been raised through the model description and experiments. First, in this work, results have been reported for a number K of parcels manually set although model selection tools based on the free-energy computation could have been used to perform this task [37] but at the expense of an increased computation load. On the one hand, a too small K may actually induce an "under-segmentation" effect where slightly different parcels are merged leading to inaccurate hemodynamic territories. On the other hand, a tool large K may induce empty parcels. The reason is that each voxel is allocated to a given parcel based on the maximum value of the $\widetilde{p}_{z_j}(k)$. If K is too large, some HRF groups will be left empty even if their HRF pattern $\bar{\boldsymbol{h}}_k$ can be estimated. This can be explained through Eq. (37) where it is shown that each voxel has its own contribution in the estimation of all HRF patterns $\bar{\boldsymbol{h}}_k$, but with different proportions depending on the value of $\widetilde{p}_{z_j}(k)$ for $k = 1 \ldots, K$. For this reason, the algorithm can end up with a number of non-empty parcels smaller than K. Second, based on a single fMRI experiment, the JPDE formalism cannot provide a robust *whole brain* parcellation into homogeneous hemodynamic territories. The reason is that evoked brain activity has to be elicited in all brain regions to enable a precise delineation of such hemodynamic parcels throughout the brain. Hence, so far the JPDE has been only helpful for studying specific brain areas involved in some specific cognitive functions (eg, language, mental computation). In the near future, owing the *individual brain charting initiative* in the Human Brain Project[2], a few individuals will be scanned several tens of times on about 50 different functional protocols. This huge dataset might be helpful for getting a whole brain hemodynamic atlas. Last, for doing so, in the meantime, the JPDE approach should be extended to perform group-level parcellation where the latter should reflect group-level hemodynamic territories.

[2]https://team.inria.fr/parietal/

VI. Conclusion

In this paper, we fully detailed and investigated the use of the JPDE framework for providing an accurate hemodynamic parcellation of a brain system (eg, the superior temporal sulcus) from fMRI data. This model removes the commitment to a priori fixed territories assumed in its JDE ancestor. By enabling a fully adaptive data-dependent identification of the parcels, the JPDE framework greatly extends the possibilities of detection-estimation approaches by allowing interaction between the detection, estimation and parcellation tasks. Obtained results on simulated and real data showed the robustness and the accuracy of the JPDE model in the three aforementioned tasks. An important pending issue remains how to optimally choose the right number of HRF groups (parcels). Nonparametric Bayesian statistics might offer the right framework to address this point. Other perspectives would be dedicated to incorporate anatomical knowledge in the HRF group prior by extending the current Potts field. Last, future work also includes the development of a population level JPDE framework in order to target a brain atlas of hemodynamics.

References

[1] L. Chaari, F. Forbes, T. Vincent, and P. Ciuciu, "Hemodynamic-informed parcellation of fMRI data in a variational joint detection estimation framework," in *MICCAI'12*, vol. 7512 of *LNCS*, pp. 180–188. Springer Berlin Heidelberg, 2012.

[2] B. Thirion, G. Flandin, P. Pinel, A. Roche, P. Ciuciu, and J.-B. Poline, "Dealing with the shortcomings of spatial normalization: Multi-subject parcellation of fMRI datasets," *Hum. Brain Mapp.*, vol. 27, no. 8, pp. 678–693, Aug. 2006.

[3] B. Thirion, P. Pinel, A. Tucholka, A. Roche, P. Ciuciu, J-F. Mangin, and J-B. Poline, "Structural analysis of fMRI data revisited: Improving the sensitivity and reliability of fMRI group studies," *IEEE Trans. Medical Imaging*, vol. 26, no. 9, pp. 1256–1269, 2007.

[4] G. Flandin, F. Kherif, X. Pennec, D. Rivière, N. Ayache, and J.-B. Poline, "A new representation of fMRI data using anatomo-functional constraints," in *Proc. 8th HBM*, Sendai, Japan, June 2002.

[5] T. Vincent, P. Ciuciu, and B. Thirion, "Sensitivity analysis of parcellation in the joint detection-estimation of brain activity in fMRI," .

[6] Y. Ji, P.-Y. Hervé, U. Aickelin, and A. Pitiot, "Parcellation of fMRI datasets with ICA and PLS–a data driven approach.," in *MICCAI'09*, G.-Z. Yang, D. J. Hawkes, D. Rueckert, J. A. Noble, and C.-J. Taylor, Eds., vol. 5761 of *LNCS*, pp. 984–991. Springer Berlin Heidelberg, Sep. 2009.

[7] S. Badillo, G. Varoquaux, and P. Ciuciu, "Hemodynamic estimation based on consensus clustering," in *IEEE PRNI'13*, Philadelphia, United States, June 2013.

[8] B. Biswal, F. Z. Yetkin, V. M. Haughton, and J. S. Hyde, "Functional connectivity in the motor cortex of resting human brain using echo-planar MRI," *Magn. Reson. Med.*, vol. 34, no. 4, pp. 537–541, Oct. 1995.

[9] V.D. Calhoun, T. Adali, G.D. Pearlson, and J.J. Pekar, "Spatial and temporal independent component analysis of functional MRI data containing a pair of task-related waveforms," *Hum. Brain Mapp.*, vol. 13, pp. 43–53, 2001.

[10] R. E. Marcus, A. M. MacLeod, A. Z. Snyder , W. J. Powers, D. A. Gusnard, and G. L. Shulman, "A default mode of brain function," *Proc. of the National Academy of Sciences*, vol. 98, no. 2, pp. 676–682, 2001.

[11] C. F. Beckmann and S. M. Smith, "Probabilistic independent component analysis for functional magnetic resonance imaging," *IEEE Trans. Medical Imaging*, vol. 23, no. 2, pp. 137–152, Feb. 2004.

[12] G. Varoquaux, S. Sadaghiani, P. Pinel, A. Kleinschmidt, J-B. Poline, and B. Thirion, "A group model for stable multi-subject ICA on fMRI datasets," *Neuroimage*, vol. 51, no. 1, pp. 288–299, 2010.

[13] G. Varoquaux, A. Gramfort, F. Pedregosa, V. Michel, and B. Thirion, "Multi-subject dictionary learning to segment an atlas of brain spontaneous activity," in *Info. Process. in Medical Imag.*, pp. 562–573. Springer Berlin Heidelberg, 2011.

[14] G. Varoquaux and R. C. Craddock, "Learning and comparing functional connectomes across subjects," *Neuroimage*, vol. 80, pp. 405–415, 2013.

[15] A. Abraham, E. Dohmatob, B. Thirion, D. Samaras, and G. Varoquaux, "Extracting brain regions from rest fMRI with total-variation constrained dictionary learning," in *MICCAI'13*, pp. 607–615. Springer Berlin Heidelberg, 2013.

[16] A. Frau, T. Vincent, F. Forbes, and P. Ciuciu, "Hemodynamically informed parcellation of cerebral fMRI data," in *IEEE ICASSP*, Florence, Italy, May 26 - 31 2014, pp. 2079 – 2083.

[17] D. A. Handwerker, J. M. Ollinger, and M. D'Esposito, "Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses," *NeuroImage*, vol. 21, pp. 1639–1651, Apr. 2004.

[18] S. Badillo, T. Vincent, and P. Ciuciu, "Group-level impacts of within- and between-subject hemodynamic variability in fMRI," *Neuroimage*, vol. 82, pp. 433–448, 2013.

[19] S. Makni, J. Idier, T. Vincent, B. Thirion, G. Dehaene-Lambertz, and P. Ciuciu, "A fully Bayesian approach to the parcel-based detection-estimation of brain activity in fMRI," *Neuroimage*, vol. 41, no. 3, pp. 941–969, 2008.

[20] T. Vincent, L. Risser, and P. Ciuciu, "Spatially adaptive mixture modeling for analysis of within-subject fMRI time series," *IEEE Trans. Medical Imaging*, vol. 29, pp. 1059–1074, 2010.

[21] L. Chaari, T. Vincent, F. Forbes, M. Dojat, and P. Ciuciu, "Fast joint detection-estimation of evoked brain activity in event-related fMRI using a variational approach," *IEEE Trans. Med. Imag.*, vol. 32, no. 5, pp. 821–837, May 2013.

[22] K. J. Friston, A. P. Holmes, K. J. Worsley, J. P. Poline, C. D. Frith, and Richard R. S. J. Frackowiak, "Statistical parametric maps in functional imaging: a general linear approach," *Hum. Brain Mapp.*, vol. 2, no. 4, pp. 189–210, 1994.

[23] L. Chaari, F. Forbes, T. Vincent, M. Dojat, and P. Ciuciu, "Variational solution to the joint detection estimation of brain activity in fMRI," in *MICCAI'11*, G. Fichtinger, A. Martel and T. Peters, Eds., vol. 6892 of *LNCS*, pp. 260–268. Springer Berlin Heidelberg, Sep. 2011.

[24] M. Woolrich, B. Ripley, M. Brady, and S. Smith, "Temporal autocorrelation in univariate linear modelling of fMRI data," *NeuroImage*, vol. 14, no. 6, pp. 1370–1386, Dec. 2001.

[25] M. Woolrich, M. Jenkinson, J. Brady, and S. Smith, "Fully Bayesian spatio-temporal modelling of fMRI data," *IEEE Trans. Med. Imag.*, vol. 23, no. 2, pp. 213–231, Feb. 2004.

[26] W.D. Penny, G. Flandin, and N. Trujillo-Bareto, "Bayesian comparison of spatially regularised general linear models," *Hum. Brain Mapp.*, vol. 28, no. 4, pp. 275–293, Apr. 2007.

[27] W. D. Penny, N. Trujillo-Barreto, and K. J. Friston, "Bayesian fMRI time series analysis with spatial priors," *NeuroImage*, vol. 23, no. 2, pp. 350–362, 2005.

[28] R.M. Neal and G.E. Hinton, "A view of the EM algorithm that justifies incremental, sparse and other variants," in *Lear. in Graph. Mod.*, Jordan, Ed., pp. 355–368. 1998.

[29] G. Celeux, F. Forbes, and N. Peyrard, "EM procedures using mean field-like approximations for Markov model-based image segmentation," *Patt. Rec.*, vol. 36, pp. 131–144, 2003.

[30] T. Sørensen, "A method of establishing groups of equal amplitude in plant sociology based on similarity of species content," *Kongelige Danske Videnskabernes Selskab*, vol. 4, pp. 1–34, 1948.

[31] L. Chaari, S. Badillo, T. Vincent, G. Dehaene-Lambertz, F. Forbes, and P. Ciuciu, "Hemodynamic-informed parcellation of fMRI data in a joint detection estimation framework," Tech. Rep., Univ. of Toulouse, Nov. 2015, https://hal.archives-ouvertes.fr/hal-01228007.

[32] K. Monzalvo, J. Fluss, C. Billard, S. Dehaene, and G. Dehaene-Lambertz, "Cortical networks for vision and language in dyslexic and normal children of variable socio-economic status," *Neuroimage*, vol. 61, no. 1, pp. 258–274, Feb. 2012.

[33] G. Dehaene-Lambertz, L. Hertz-Pannier, J. Dubois, S. Mériaux, A. Roche, and M. Sigman, "Functional organization of perisylvian activation during presentation of sentences in preverbal infants," *Proc. Natl. Acad. Sci. USA*, vol. 103, pp. 14240–14245, 2006.

[34] J. Brauer, J. Neumann, and A. D. Friederici, "Temporal dynamics of perisylvian activation during language processing in children and adults," *Neuroimage*, vol. 41, no. 4, pp. 1484–1492, 2009.

[35] G. Dehaene-Lambertz, L. Hertz-Pannier, and J. Dubois, "Nature and nurture in language acquisition: anatomical and functional brain-imaging studies in infants," *Trends in Neurosci.*, vol. 29, pp. 367–373, 2006.

[36] P. Ciuciu, S. Sockeel, T. Vincent, and J. Idier, "Modelling the neurovascular habituation effect on fMRI time series," in *IEEE ICASSP*, Taipei, Taiwan, Apr. 2009, pp. 433–436.

[37] M. Albughdadi, L. Chaari, F. Forbes, J.-Y. Tourneret, and P. Ciuciu, "Model selection for hemodynamic brain parcellation in fMRI," in *22nd EUSIPCO*, Lisbon, Portugal, Sep. 2014, pp. 31–35.